# Exploiting Conditional Independence in Ensemble Data Assimilation

Patricia Kappler, Department of Geoscience & Engineering, TU Delft

## 1 | OVERVIEW DATA ASSIMILATION

### Data Assimilation (DA)

DA combines numerical modelling with observed data to improve parameter and state estimation. DA is based on **Bayes' theorem** which combines a **prior pdf** (representing prior knowledge) and a **likelihood pdf** (introducing observed data) to estimate the **posterior pdf**.

**Bayes' theorem:**

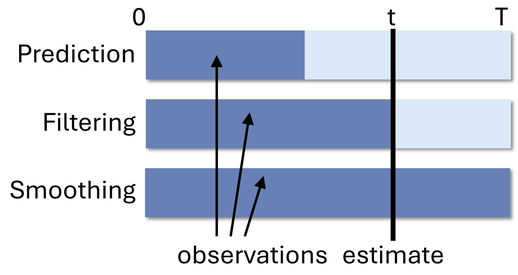
$$p(a|b) = \frac{p(b|a)p(a)}{p(b)}$$

posterior: $p(a|b)$; likelihood: $p(b|a)$; prior: $p(a)$; normalizing factor: $p(b)$

### Ensemble Methods

For DA methods is the covariance matrix essential. Ensemble methods derive the covariance matrix from a finite ensemble as the true closed-form solution to DA's statistical inference problems is often intractable.

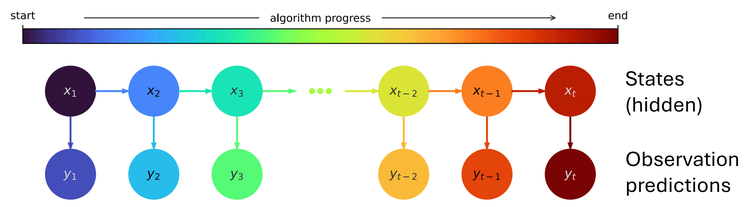
### Prediction, Filtering, Smoothing




Adapted from Särkkä, Svensson (2023).

DA can be used to solve different state estimation problems depending on the time span of available observations.

### Conditional Independence




Hidden Markov model. Adapted from Ramgraber et al. (2023).

Pdfs have an underlying graphical structure, which represents variables as nodes and dependencies as edges and encodes conditional independence (CI): $A \perp B|C$ if all paths from $A$ to $B$ go through $C$. **Example**: In a hidden Markov model the current state is independent of past observations given the previous state: $x_t \perp y_{t-1}|x_{t-1}$.

## 2 | OBJECTIVE

Small ensemble sizes lead to spurious correlations, which can be reduced by exploiting CI. Some DA methods **exploit CI** (in time), **some not**:

- Ensemble Kalman Filter (EnKF) (CI)
- Bulk Ensemble Kalman Smoother (Bulk EnKS) (not CI)
- Ensemble Rauch-Tung-Striebel Smoother (EnRTSS) (CI)

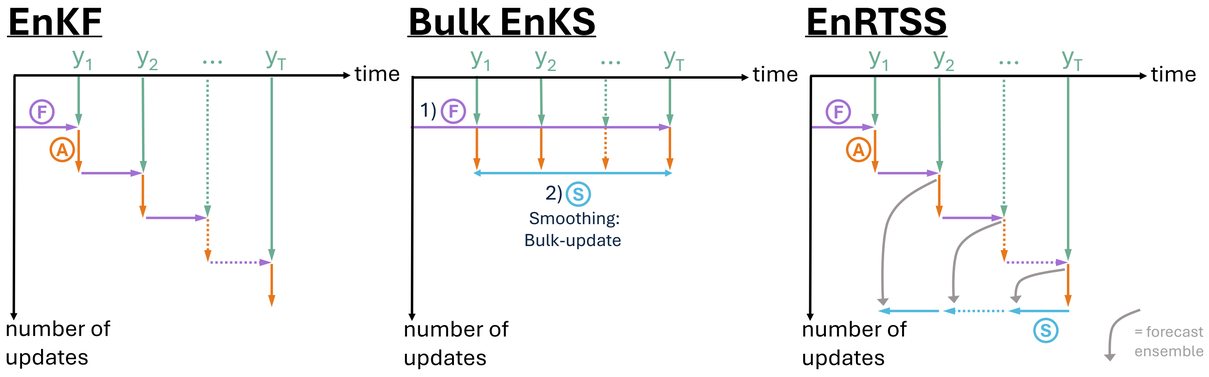
**Model:** first order autoregressive model (AR(1)) $x_t = \phi x_{t-1} + \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2)$

## 3 | MATHEMATICAL FOUNDATIONS

All Kalman-type DA algorithms use Bayesian inference. They differ in what they update and what they condition on:

1. Form joint distribution: $p(a, b|c)$
2. Assume $p(a, b|c)$ is Gaussian
3. Condition $b$ on $b^*$ (specific observation) to get $p(a|b^*, c)$
4. Estimate mean and covariance from samples
5. Update samples: $A^* = A - \Sigma_{a,b}\Sigma_{b,b}^{-1}(B - b^*)$
   ($A$ and $B$ are the ensemble matrices for $a$ and $b$)

EnKF and EnRTSS are recursive and make use of CI in time. The bulk EnKS performs a bulk update and is not recursive.




F Forecast  A Analysis/Update  S Smoothing  $t = 1,2,..T$

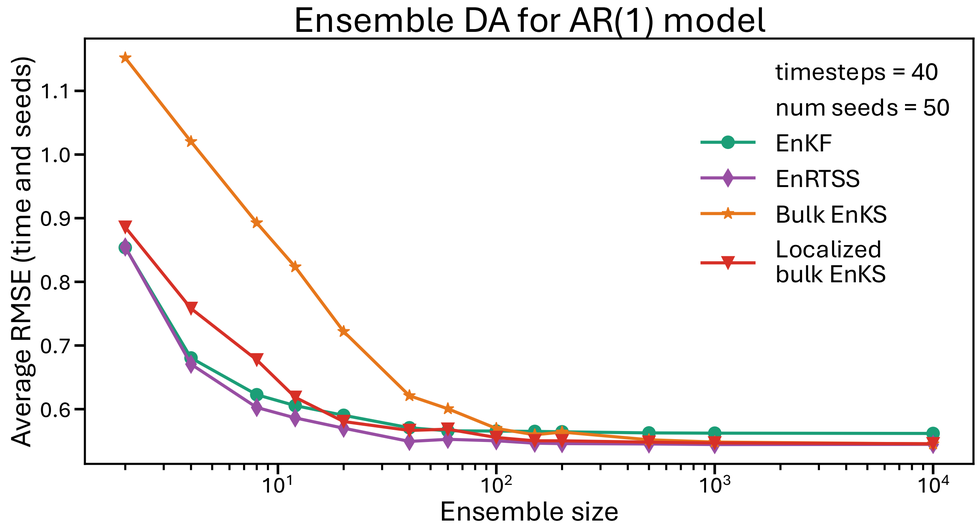
$a = x_t^f, b = y_t, c = y_{1:t-1}^*, b^* = y_t^*$

$a = x_{1:T}, b = y_{1:T}, c = \emptyset, b^* = y_{1:T}^*$

$a = x_t^a, b = x_{t+1}^f, c = x_{1:t}^*, b^* = x_{t+1}^s$

$x_t^f \sim p(x_t|y_{1:t-1}), \quad x_t^a \sim p(x_t|y_{1:t}), \quad x_t^s \sim p(x_t|y_{1:T})$

## 4 | RESULTS




**Small ensemble sizes:**

EnRTSS > EnKF > EnKS

estimation accuracy

- EnKF and EnRTSS exploit CI in time; bulk EnKS not
- Localization of bulk EnKS improves accuracy
- Theoretically, smoothers outperform EnKF by using future observations. **But:** EnKS is prone to spurious correlations

**Large ensemble sizes:**

EnRTSS (loc) EnKS > EnKF

estimation accuracy

- Covariance approximation of EnKS and EnRTSS become more accurate: both smoothers yield similar accuracy
- All smoothers perform better than EnKF: Fewer spurious correlations for large ensembles

The ranking of methods by estimation accuracy is in line with theory. **Exploiting CI** in ensemble DA **improves** estimation accuracy for **small ensembles**, while the impact diminishes for large ensembles.

## 5 | OUTLOOK

- Geosciences → small ensemble size → exploiting CI in time useful
- CI not only in time but also in space
- Challenge: Finding CI in space – How?


**References:**

Särkkä, S.; Svensson, L. Bayesian Filtering and Smoothing, Second edition.; Institute of Mathematical Statistics textbooks; Cambridge University Press: New York, 2023.

Ramgraber, M.; Baptista, R.; McLaughlin, D.; Marzouk, Y. Ensemble Transport Smoothing. Part I: Unified Framework. J. Comput. Phys.: X 2023, 17, 100134. https://doi.org/10.1016/j.jcpx.2023.100134.

Evensen, G. Data Assimilation: The Ensemble Kalman Filter; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009. https://doi.org/10.1007/978-3-642-03711-5.

Cosme, E.; Verron, J.; Brasseur, P.; Blum, J.; Auroux, D. Smoothing Problems in a Bayesian Framework and Their Linear Gaussian Solutions. Mon. Wea. Rev. 2012, 140 (2), 683–695. https://doi.org/10.1175/MWR-D-10-05025.1.

Evensen, G.; Van Leeuwen, P. J. An Ensemble Kalman Smoother for Nonlinear Dynamics. Mon. Wea. Rev. 2000, 128 (6), 1852–1867. https://doi.org/10.1175/1520-0493(2000)128%253C1852:AEKSFN%253E2.0.CO;2.

TU Delft